# DIRECT TESTIMONY

# OF

# R. DOW BAILEY

# ON BEHALF OF

# SOUTH CAROLINA ELECTRIC & GAS COMPANY

# DOCKET NO. 2004-2-E

**Q. PLEASE STATE YOUR NAME AND BUSINESS ADDRESS.**

**A.** R. Dow Bailey, 1426 Main Street, Columbia, South Carolina.

**Q. BY WHOM ARE YOU EMPLOYED AND IN WHAT CAPACITY?**

**A.** I am Forecast Coordinator in the Resource Planning Department of SCANA Corporation.

**Q. PLEASE DESCRIBE YOUR EDUCATIONAL BACKGROUND AND BUSINESS EXPERIENCE.**

**A.** I am a graduate of Emory University in Atlanta, Georgia where I majored in history. I also received an MBA from the University of Georgia, with an emphasis on finance and economics. I have completed all the coursework requirements for a Ph.D. in economics at the University of South Carolina. In addition to these academic studies I have attended numerous seminars on forecasting and statistics, sponsored by such organizations as NARUC, DOE, the Electric Power Research Institute (EPRI), the National Association of Business Economists (NABE), and the American Gas Association (AGA). Prior to my employment with SCE&G I was employed as an Economic Analyst with Gulf Oil Corporation; an Economist with Wilbur Smith & Associates; a Research Analyst with the South Carolina Public Service Commission; an Economist with CH2M Hill, a consulting engineering firm; and a Financial Analyst with Northeast Utilities. In June 1983 I began work at SCE&G as an Associate Analyst in the Forecasting Department, where I have been employed for the past nineteen years.

**Q. WILL YOU BRIEFLY SUMMARIZE YOUR DUTIES WITH SOUTH CAROLINA ELECTRIC & GAS COMPANY?**

**A.** I am currently responsible for preparing SCE&G's electric and gas forecasts of sales, customers, revenues, and peak demand , as well as other forecasting duties within SCANA.

**Q. WHAT IS THE PURPOSE OF YOUR TESTIMONY?**

**A.** The purpose of my testimony is to discuss the construction of the short-run electric sales forecast. Short-run is defined as two years, in this case 2004 and 2005.

**Q. PLEASE DESCRIBE HOW THE SALES FORECAST WAS DETERMINED.**

**A.** The first step in the process was to develop a data set containing monthly information on SCE&G customers, sales, and average use at a class and/or rate level. A number of large industrial customers were separated into individual accounts, as were municipal and cooperative customers. The size of these customers justified developing projections for them individually. Within the residential sector, customers were grouped into three major categories: single-family, multi-family, and mobile home, and further disaggregated into space and non-space heat groups . This is important because there are distinct differences in usage within these housing and heating types that can only be modeled correctly by separating them from each other.

At the same time daily weather data from Charleston and Columbia was also gathered for the same historic period. The bulk of SCE&G's residential and commercial customers are located in these two cities, so using an average of the two makes estimation of weather-related sales more reliable. The daily sales were then weighted by scheduled billing cycles to create cycle-weighted weather variables. Weather data calculated on a calendar basis does not correlate well with cycle-billed sales, because the latter are spread over two months, rather than one, so this problem is solved by the weighting process. Weather was represented by heating degree days (HDD) or cooling degree days (CDD). HDD and CDD are defined as the absolute value of the difference between the average daily temperature for a given weather station and a base temperature. HDD are zero if the average daily temperature is above the base temperature, with the converse being true for CDD. Positive HDD indicate the need for space heating and positive CDD likewise the need for space cooling. At SCE&G, based on historical experience, the CDD and HDD bases were set at 60° and 75°, respectively. Depending on the season, summer or winter, either CDD or HDD were used to represent weather in the forecast. The months of November through April were considered winter months, with May through October defined as summer. After assembling the data, the actual models were developed. Exhibit__(RDB-1) shows the class/rate breakdown used in the forecast. There are 31 class/rate groupings. However, the use of seasonal and individual customer models means that more equations were developed than the number of classes and rates shown in the exhibit. Overall, nearly 100 models were utilized to develop the short-run forecast.

**Q. WHAT TYPES OF MODELS WERE USED TO DEVELOP THE FORECAST?**

**A.** Two different types of statistically based models were used to create the forecast, time-series and regressions. Time-series models, from the class of models sometimes known as "Box-Jenkins or "ARIMA" models, were developed to project the bulk of the customer projections. Some customer groups, e.g., municipals and cooperatives, had their customer levels set at a fixed amount because the addition of this type of account is generally known well in advance. On the other hand, classes with large numbers of accounts, such as the residential group, increase according to broad economic and demographic trends within SCE&G's service area. In the short-run, these movements are captured quite accurately by time-series models. In a very few cases time-series models were also used to forecast sales or average use for customers whose usage patterns were not very sensitive to weather.

Regression models are a means of statistically relating one variable, such as average use, to other "explanatory" variables, such as weather, price, industrial output, etc. To create a forecast via regression models, first an estimated equation is created using historical data. A number of different model specifications are tried before the final equation is selected. Choice of the final equation is based upon the equation's statistical characteristics, such as its consistency with actual data and the statistical significance of model parameters, its ability to capture turning points in the data, and the reasonableness of equation results. The chosen model is then combined with projected values of the explanatory values to derive the forecast of the dependent variable. If the dependent variable was average use for a given category, as was generally the case, this value was then combined with projected customers to derive sales. Exhibit__(RDB-2) lists the methodologies used to forecast the major rate/class groupings.

**Q. CAN YOU GIVE AN EXAMPLE OF AN EQUATION USED TO DEVELOP THE FORECAST?**

**A.** Yes. Exhibit__(RDB-3) shows a graph relating summer residential average use for the single-family space-heating group to CDD, while Exhibit__(RDB-4) contains output describing regression model results for this same category. The graph in Exhibit__(RDB-3) was included because it illustrates the fact that average use seems more responsive to CDD at lower values than at higher values, a somewhat counter-intuitive result. This phenomenon occurs primarily because lower CDD occur in May and October, the beginning and end of the summer season. In May this result can be explained by startup of air-conditioning units by customers, which when aggregated over a group gives rise to a higher than normal response. In October the reverse seems to be true, and at any rate this factor needs to be accounted for in the forecast. To do so CDD were separated into two groups, CDDLO and CDDHI, so separate coefficients could be estimated. These values, shown under the column heading "Parameter Estimates", were 5.40 and 3.05, respectively. The interpretation of the 5.40 coefficient is that average use will increase by 5.4KWH for each CDD, as long as the total CDD are less than 100. When CDD are greater than 100, then the response drops to 3.05 KWH for each CDD.

**Q. WHAT DO THE OTHER VARIABLES IN THE EQUATION INDICATE?**

**A.** There are five variables in the equation which begin with "Y" followed by a year and month, e.g., Y8909. These are sometimes called dummy, indicator or binary variables, and are used to account for the impact of unusual events or other factors not explained by standard explanatory variables. For example, the value of –240.04 was estimated by the regression for Y8909. This represents an estimate of reduced average use during the month of September 1989, when Hurricane Hugo occurred. Accounting for this unusual event allows the other variables to be more accurately estimated in the regression process.

SEP is another dummy variable which provides an estimate of higher average use in the month of September after the effects of the other explanatory variables are taken into consideration. September is a transition month, but due to cycle billing a large amount of usage from August is actually recorded in September. Consequently, average use for the month is higher than predicted by a model relying solely on weather, and this fact is taken into consideration with the SEP dummy variable. The final variable, TIME, is used to estimate a value for increased average usage solely as a function of time among this group of customers. There are a number of likely explanations for this fact, among them increasing incomes, larger homes, more appliances, or the decreasing real price of electricity. TIME is used as a proxy of their cumulative effects.

The columns headed by "t Value" and "Pr > |t|" indicate the statistical significance of the variables used in the regression equation. Generally speaking, t values greater in absolute value than 2 indicate that a variable is statistically significant. All of the variables used in this equation meet that criteria.

**Q. WHAT OTHER STATISTICS IN EXHIBIT (RDB-4) ARE USED TO JUDGE THE ADEQUACY OF THE EQUATION FOR FORECASTING?**

A. Several of the statistics contained in the Exhibit (RDB-4) regression output point to its suitability as a method to forecast electricity sales. One of the most commonly examined of these "goodness of fit" indicators is the adjusted $R^2$, which is shown there as "Adj R-Sq". $R^2$ ranges from zero to one, with values close to one indicating better capability of a regression equation to explain movements in the dependent variable. The value here is 0.9872, which means that almost 99% of the variation in average use for this group is explained by the regression equation. The statistic "Root MSE" is the root mean square error of the regression. It measures the average error of the model, so when comparing competing specifications the equation with the lower of the two is considered more accurate. Here the value of 37.87 can be compared with the mean of monthly average use, which is calculated to be 1585.55. Thus the unexplained variability, or average error of the equation, is about 2.4%. A number of other statistics and graphs not shown in the exhibit were also used in the decision to use this particular specification of residential average use.

**Q. WHAT IS THE NEXT STEP IN THE FORECAST PROCESS?**

**A.** After all the statistical work is completed and the models have been used to generate a preliminary forecast, the projections are then compared to historic actuals for reasonableness. If certain events affecting sales or customer growth in the future are known, but not contained in the model structure, adjustments are also made for these factors. A prime example of this is information gathered by the Industrial Marketing and Economic Development departments on SCE&G's large customers. Typically they will know if a customer is planning to expand its operations or if a new large customer is requesting service. Finally, when the overall territorial sales forecast is assembled it is reviewed once more to see if it is logical and in line with expectations.

**Q. HOW ACCURATE HAS THE SHORT RANGE FORECAST METHODOLGY BEEN AT SCE&G?**

**A.** Over the past ten years, 1994-2003, the mean absolute percentage error (MAPE) of the short range forecast was just over 1% comparing actual weather-normalized sales to the projections developed for that year's budget. The largest error of –2.2% occurred in 2001, and was primarily due to the economic slowdown which took place in the second half of the year, while the smallest error was –0.2% in 1996. In 2003 there was a difference of –0.5%.

**Q. DOES THIS CONCLUDE YOUR TESTIMONY?**

**A.** Yes.

Short-Term Forecasting Groups, 2004 – 2005

| Class Number | Class Name | Rate/SIC Designation | Comment |
|---|---|---|---|
| | | Single Family | Rates 1, 2, 5, 6, 8, 18, 25, 26, 62, 64 |
| 10 | Residential Non-Space Heating | Multi Family | Rates 67, 68, 69 |
| 910 | Residential Space Heating | Mobile Homes | Rates 1, 2, 5, 7, 8 |
| 20 | Commercial Non-Space Heating | Rate 9 | Small General Service |
| | | Rate 12 | Churches |
| | | Rate 20, 21 | Medium General Service |
| | | Rate 22 | Schools |
| | | Rate 24 | Large General Service |
| | | Other | Rates 10, 11, 14, 16, 17, 18, 24, 25, 26, 29, 60, 62, 64, 67, 68, 69 |
| 920 | Commercial Space Heating | Rate 9 | Small General Service |
| 30 | Industrial Non-Space Heating | Rate 9 | Small General Service |
| | | Rate 20, 21 | Medium General Service |
| | | Rate 23, SIC 22 | Textile Mill Products |
| | | Rate 23, SIC 24 | Lumber, Wood Products,Furniture and Fixtures (SIC Codes 24 and 25) |
| | | Rate 23, SIC 26 | Paper and Allied Products |
| | | Rate 23, SIC 28 | Chemical and Allied Products |
| | | Rate 23, SIC 30 | Rubber and Miscellaneous Products |
| | | Rate 23, SIC 32 | Stone, Clay, Glass, and Concrete |
| | | Rate 23, SIC 33 | Primary Metal Industries; Fabricated Metal Products; Machinery; Electric and Electronic Machinery, Equipment and Supplies; and Transportation Equipment (SIC Codes 33-37) |
| | | Rate 23, SIC 91 | Executive, Legislative and General Government (except Finance) |
| | | Rate 23, SIC 99 | Other or Unknown SIC Code* |
| | | Rate 27, 60 | Large General Service |
| | | Other | Rates 25 and 26 |
| 930 | Industrial Space Heating | Rate 9 | Small General Service |
| 60 | Street Lighting | Rates 3, 9, 13, 17, 25, 26, 29, and 69 | |
| 70 | Other Public Authority | Rate 3 and 29 | |
| | | Rates 65 and 66 | |
| 92 | Municipal | Rate 60, 61 | Four Individual Accounts |
| 97 | Cooperative | Rate 60, 61 | Four Individual Accounts |

Includes small industrial customers from all SIC classifications that were not previously forecasted individually.

Note: Industrial Rate 23 also includes Rate 24. Commercial Rate 24 also includes Rate 23.

Summary of Methodologies Used To Produce
2004 and 2005 Short Range Forecast

| Value Forecasted | Methodology | Forecasting Groups |
|---|---|---|
| Average Use | Regression | Class 10, All Groups<br>Class 910, All Groups<br>Class 20, Rates 9, 12, 20, 22, 24, 99<br>Class 920, Rate 9<br>Class 70, Rate 3 |
| Total Usage | ARIMA/<br>Regression | Class 30, Rates 9, 20, 99, and 23,<br>for SIC = 91 and 99<br>Class 930, Rate 9<br>Class 60<br>Class 70, Rates 65, 66 |
| | Regression | Class 92, All Accounts<br>Class 97, All Accounts |
| Customers | ARIMA | Class 10, All Groups<br>Class 910, All Groups<br>Class 20, All Rates<br>Class 920, Rate 9<br>Class 30, All Rates Except 60, 99, and 23<br>for SIC = 22, 24, 26, 28, 30, 32, 33, and 91<br>Class 930, Rate 9<br>Class 60<br>Class 70, Rate 3 |

MAY - OCTOBER
SINGLE FAMILY HOMES SPACE HEATING

Plot of AVG*CDD. Symbol is value of mo.

```
 AVG ,
2200 ^                                            87
     ,
     ,                                    7 7
     ,                                  8
     ,                              9         8
2000 ^                                78        8
     ,                           7     88
     ,                           7
     ,                        9  89  7
     ,                    9  9 88  7
1800 ^                       6 87 8   8
     ,
     ,             9     7 9
     ,                 9
     ,              9 9
1600 ^             6 6 66   6
     ,           6 96  6       9
     ,
     ,      6
     ,             6
1400 ^
     ,    5
     ,    61 11
     ,
     ,  151
1200 ^ 5 5  115 1
     ,  11 5 15
     ,   15
     , 5555
     ,  5
1000 ^
     Šff^fffffffffffff^fffffffffffff^fffffffffffff^fffffffffffff^fffffffffffff^fffffffffffff^fffffffffffff^fffffffffffff^ff
       0.0     50.0     100.0     150.0     200.0     250.0     300.0     350.0

                                 CDD
```

MAY - OCTOBER
SINGLE FAMILY HOMES SPACE HEATING

The REG Procedure
Model: MODEL1
Dependent Variable: AVG

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 9 | 8764686 | 973854 | 678.99 | <.0001 |
| Error | 70 | 100398 | 1434.26232 | | |
| Corrected Total | 79 | 8865084 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 37.87166 | R-Square | 0.9887 |
| Dependent Mean | 1585.54507 | Adj R-Sq | 0.9872 |
| Coeff Var | 2.38856 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Type I SS | Type II SS |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 1012.26153 | 12.80657 | 79.04 | <.0001 | 201116253 | 8960829 |
| CDDLO | 1 | 5.40359 | 0.17370 | 31.11 | <.0001 | 7228979 | 1387962 |
| CDDHI | 1 | 3.04862 | 0.10465 | 29.13 | <.0001 | 1356334 | 1217208 |
| Y8909 | 1 | -240.03647 | 40.26384 | -5.96 | <.0001 | 62178 | 50974 |
| Y8907 | 1 | 131.79419 | 39.06228 | 3.37 | 0.0012 | 7965.51475 | 16327 |
| Y9406 | 1 | -128.78835 | 39.40596 | -3.27 | 0.0017 | 24314 | 15320 |
| Y9609 | 1 | 83.92447 | 39.88266 | 2.10 | 0.0389 | 10327 | 6350.93284 |
| Y990506 | 1 | 108.90151 | 27.91055 | 3.90 | 0.0002 | 30573 | 21835 |
| SEP | 1 | 32.12583 | 13.57617 | 2.37 | 0.0207 | 9020.76883 | 8031.24794 |
| TIME | 1 | 0.95232 | 0.19279 | 4.94 | <.0001 | 34995 | 34995 |

| | |
|---|---|
| Durbin-Watson D | 1.894 |
| Number of Observations | 80 |
| 1st Order Autocorrelation | 0.032 |